DOCUMENT RESUME

ED 037 827 24 EA 002 819

AUTHOR Scriven, Michael
TITLE Evaluation as a Main Aim of Science: Comments on Professor Bloom's Paper Entitled "Toward a Theory of Testing Which Includes Measurement-Evaluation-Assessment."
INSTITUTION California Univ., Los Angeles. Center for the Study of Evaluation.
SPONS AGENCY Office of Education (DHEW), Washington, D.C. Bureau of Research.
REPORT NO CSE-R-10
BUREAU NO BR-6-1646
PUB DATE Sep 69
CONTRACT OEC-4-6-061646-1909
NOTE 12p.; From the Proceedings of the Symposium on Problems in the Evaluation of Instruction (Los Angeles, December 1967)

EDRS PRICE EDRS Price MF-$0.25 HC-$0.70
DESCRIPTORS *Evaluation, *Measurement, Scientific Attitudes, *Scientific Methodology, *Testing, *Values

ABSTRACT

Both measurement and testing are regarded as component methodologies contributing to the legitimate and more comprehensive scientific activity of evaluation. This perspective is broader than that offered by Bloom and incorporates the need for the educational evaluator not only to apply specific criteria and methods for evaluation but also to exercise value judgments determining their appropriateness. A related document is EA 002 818. (JK)

ED037827

BR 6-1646
PA-24
OE/BR
EA

CSE Report No. 10

EVALUATION AS A MAIN AIM OF SCIENCE: COMMENTS ON PROFESSOR BLOOM'S PAPER ENTITLED "TOWARD A THEORY OF TESTING WHICH INCLUDES MEASUREMENT-EVALUATION-ASSESSMENT"

Michael Scriven

EA 002 819

Center FOR THE
Study of
Evaluation
OF INSTRUCTIONAL
PROGRAMS

University of California, Los Angeles, September 1969

CO-DIRECTORS

Merlin C. Wittrock Erick L. Lindman

ASSOCIATE DIRECTORS

Marvin C. Alkin Frank Massey, Jr. C. Robert Pace

The CENTER FOR THE STUDY OF EVALUATION OF INSTRUCTIONAL PROGRAMS is engaged in research that will yield new ideas and new tools capable of analyzing and evaluating instruction. Staff members are creating new ways to evaluate content of curricula, methods of teaching and the multiple effects of both on students. The CENTER is unique because of its access to Southern California's elementary, secondary and higher schools of diverse socio-economic levels and cultural backgrounds.

ED037827

EVALUATION AS A MAIN AIM OF SCIENCE: COMMENTS ON PROFESSOR BLOOM'S PAPER ENTITLED "TOWARD A THEORY OF TESTING WHICH INCLUDES MEASUREMENT-EVALUATION-ASSESSMENT"

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

Michael Scriven

University of California, Berkeley

From the Proceedings of the

SYMPOSIUM ON PROBLEMS IN THE EVALUATION OF INSTRUCTION

University of California, Los Angeles
December, 1967

M. C. Wittrock, Chairman

Sponsored by the Center for the
Study of Evaluation of Instructional Programs

The research and development reported herein was performed pursuant to a contract with the United States Department of Health, Education, and Welfare, Office of Education under the provisions of the Cooperative Research Program.

CSEIP Occasional Report No. 10, September 1968
University of California, Los Angeles

EVALUATION AS A MAIN AIM OF SCIENCE
COMMENTS ON PROFESSOR BLOOM'S PAPER ENTITLED "TOWARD A THEORY OF TESTING WHICH INCLUDES MEASUREMENT-EVALUATION-ASSESSMENT"

Michael Scriven

Introduction

The distinctions proposed by Professor Bloom are not entirely clear to me, but it is clear that an immense variety of tasks fall under each of his headings. He calls for a synthesis; I am more struck by the need for recognizing the right to autonomy. But a debate at this level of generality is unlikely to be fruitful, so let me be more specific. I think the simplifying urge in all scientists, appropriate though it is in certain areas where Occam's razor is needed to clear the jungle, is often extended into regions where we need, instead, to be tenderly nurturing new growth. Running amuck with Occam's razor can happen at the metalevel, too. In particular, an arid emphasis on the descriptive aim of science has weakened our efforts in other equally legitimate directions.

The Narrow View

A common view among empiricists is that science has only one legitimate aim--to "tell it like it is," to describe the universe in terms of general laws and particular facts. The idea that understanding involves more than description is regarded as naive psychologism; prediction is regarded as simply instantiation of descriptive laws and, hence, not significantly distinct; control is the business of engineers and politicians who may or

may not use science, but whose administrative decisions are certainly not its affair; and evaluation is pretty close to an obscenity, symptomatic of failure to achieve the value-free ideal of science.

I want to express, though I cannot defend in detail a strongly different view which I hope will allow a broader perspective of Professor Bloom's interesting paper. Curiously enough, methodological reductionists of the value-free school have been exceptionally careless in their analysis of what is for them the fundamental process, that of description. They have overlooked the highly sophisticated taxonomical developments that are presupposed by good scientific description, the differences between observation, identification, classification and description, and many other points that make scientific description of particular facts a very sophisticated process. When we turn to "general descriptions," i.e., laws, the failure to analyze is truly astonishing. It is still common to think that the logical form of scientific law is "All A's are B's."

Scope of Science

Science is an activity as well as a body of knowledge. It can be subdivided topically (into astronomy, biology, chemistry... zoology) or methodologically. The latter subdivisions, with which we are concerned, will be worth noting if they reflect substantial differences in training, practical procedures, assumptions, or instrumentation, even if there is some philosophical sense in which some of them can be reduced to others, or all to one.

Some proposed methodological subdivisions of the scientific process are: explanation, prediction, control, description, evaluation, experimentation, observation, measurement, classification, identification, diagnosis, retrodiction, analysis, generalization, deduction, induction, intuition, understanding, reflection, speculation, checking, construction of artifacts, and training. In these terms, I see the scientific aspects of the educational enterprise, for example, as (a) observation of student and other behavior (including attitudes, capacities, etc.) in putative learning situations; (b) development and application of new scales, taxonomies, generalizations, and theories to describe this behavior; (c) experimentation and theorizing to identify and distinguish the environmental and subject variables (genetic and learned) which produce the behavior; (d) evaluation of the student, his performance and the effect of the environment (including both contrived and natural features, e.g., textbooks and weather) with respect to various educational criteria [using (a) through (c)]; (e) prediction/retrodiction of later/earlier behavior, etc. [using (a) through (c)]; (f) creation of new materials to achieve desired effects [using (a) through (c)].

This account does not, I think, involve any new and potentially confusing senses of old terms like "assessment" or distinctions, which seem artificial, between it and "evaluation." It does involve a good many more distinctions than his primary three, but they are ones with which we are familiar enough in practice and seem to have more simplicity of reference. "Testing," for example

is to be seen as sometimes measurement leading to description, sometimes as checking the consequences of a theory, and sometimes as the key element in evaluation. I prefer not to define it in the totally comprehensive way that Bloom does, where it includes all data-gathering; surely field study observation is not usefully described as testing.

In the end, Bloom has far more distinctions than I do, but mine are already in the language and do not have to be learned. That is no excuse for not learning his if they really pull their weight, but there I confess that the point of many of them eludes me. So I am proposing a more conservative and less technical taxonomy than his for the scientific investigation of education, but still one that considerably expands the commonly used range of distinctions.

Scientific Evaluation

The main aim of this note, however, is to argue for a more honest and powerful conception of evaluation. "Telling it like it is," i.e., meaningful description, must involve evaluation. Owing to rather widespread methodological sloppiness and a degree of understandable cowardice by social scientists, including those in the educational field, we have diluted the threatening notion of evaluation until it has become almost as harmless as instruction or observation and hard to distinguish from testing. Not many people feel they are beyond reproach, and not very many like to administer reproach; so we have an almost united front of negative motivation about evaluation. But evaluation is an obligation that cannot be avoided by wishing it would go away.

The simple fact is that evaluation is one of the absolutely fundamental tasks and obligations of science, both pure and applied, whether or not one accepts the view that it is part of effective description. Not only is the goal of a value-free science unattainable in practice, it is absurd in principle. The aim of pure science is to produce good or better explanations, principles, and classifications (for example), and the use of the valuational terms "good" and "better" in that claim is essential. Nor can they be translated into other terms in any general way. What counts as good is different in the three cases mentioned, just as the practical criteria for a good FM tuner are different from those for a good skinning knife; but what is common is the process of evaluation and the production of value judgments, and a scientist who is not skilled in this process and production cannot be a good scientist. Even if a wholly general but still usefully applicable translation could be given for the value vocabulary, this would not banish evaluation; it would only rechristen it. There remain to be discussed three evasive moves and modifications in the methodological game of decontaminating science from value judgments.

Decontaminating Attempts

First, it is sometimes argued that this kind of evaluation is "only instrumental (or derived) evaluation," using criteria that are not established by science: the forbidden enterprise is "fundamental (or basic) evaluation," the setting up of the ultimate values. Second, it is argued that these methodological value judgments are not at issue; it is only moral value judgments

that science cannot make. Third, it may be said that in certain applied areas, like education, the relevant value criteria in fact and properly come from outside science.

It is worth summarizing the major objections to these moves, because evaluators need a philosophical platform if they are to withstand the slings and arrows of outraged peers, parents, and pupils.

It is true there is a difference between instrumental and fundamental evaluation. Judging the merits of scientific contributions, e.g., Bloom's work, is indeed different from deciding what criteria for merit are to be (should be) applied. But, I would argue that the scientific enterprise embraces both.

The search for new paradigms of theory in science often involves decisions on the criteria for a good theory--for example, is it essential that it be prediction-generating or deterministic? Surely scientists, with some help and hindrance from philosophers of science, do discuss and answer these questions in a rational manner. The conceptual analysis of the nature of science involved is a meta-scientific rather than an intrascientific task, but it is not a field where only arbitrary choices are possible. It seems entirely appropriate that metascience, especially since its results crucially affect specific sciences, should be regarded as a legitimate branch of science just as the foundations of math is regarded as a branch of math. Even if some kind of distinction of degree is made, it is certainly part of the province of the empirical and logical analysis of scientific activity, not a mere matter of taste.

Suppose someone said: in an important sense, what we define as "science" is still an arbitrary decision, and it is only that decision which makes possible the derived value judgments about the merits of theories, etc., which do admittedly occur within science. For we might just say that "science" includes theology and astrology and Christian Science, in which case the aspects of scientific theories that are to count as meritorious would also have to be redefined. So the judgments of merit depend on a lexicographical decision, which is of course arbitrary. And arbitrary decisions are not the business of science. Then one can reply that in this sense of "arbitrary," it is arbitrary that truth is not falsehood, one is not two, heat is not temperature, etc. Conversely, there is nothing less arbitrary than definitional truths, and it is a definitional truth that scientific theories should imply truths about the world. It is such definitional truths which provide the basis for methodological value judgments. The only grain of truth in the value-free line is the fact that the basic value judgments might be said to be part of the logic of science rather than of particular sciences. The fact remains that they are securely established and provide the foundation for all standards of quality in science.

The second move concedes the role of methodological value judgments in science and only excludes moral ones. In this case, we may take moral value judgments as those made about human behavior, attitudes, etc., with respect to their effects on other humans, judged from the point of view which treats humans as having prima facie equal rights. The toughest countermove here consists in

saying that ethics itself is a (social) science and that it certainly involves moral value judgments. One cannot argue against this by invoking the normative/descriptive distinction because of the previous arguments. Nor do I think it can be defeated by the usual antiutilitarian arguments about morality being an end in itself and not a matter of calculating consequences for people, since these can be met (see "Morality" in Primary Philosophy, McGraw-Hill, 1966).

An independent but weaker counter-argument suggests that in certain applied sciences moral considerations enter essentially. For example, in psychotherapy, the definition of improvement in the patient or client cannot legitimately exclude considerations of his effects on the welfare of others. This is a more powerful claim than the general observation that a scientists's output, whether bombs or banana bread, can be used for good or ill; the latter point shows only that he cannot as a person avoid some responsibility for foreseeable applications of his work. The present point suggests that his work as a scientist necessarily involves moral value judgments. The best counter is to attempt to separate the moral from the medical criteria; it is not a promising one. Even psychoanalytical theory, for all the moral relativism espoused by many of its practitioners, embodies an account of the good life and of obligations and duties. The same is true of most personality theory and approaches to educational psychology, abnormal psychology, the psychology of sex, etc. At the very least, there are moral obligations on the scientist in these areas which intermingle with scientific considerations in his work.

The final skirmish particularly concerns "service" fields like educational evaluation. Now it is true that an applied scientist in principle can evaluate almost anything with respect to almost any set of criteria. But his task as a scientist is not confined to such factual inquiries, though the exigencies of his employment situation may require it. A good research man helps in the development of evaluation criteria at both the abstract and the operational levels, whether he is concerned with detergents, hop-pickers or Headstart. He is usally not the only source of, though he may well be the chief assessor of, considerations relevant to criterion-picking, since there is usually a market involved in one sense or another. In education this is the population of parents, prospective employers, citizens, and especially the students themselves. Their needs and wants are indeed relevant, but none of these groups has absolute priority over the others; nor are any of them well-equipped to identify their unfelt needs, to translate their felt needs into educationally usable criteria, or to give an intelligent weighting to possible side-effects. For this we need the skills of social science. So the educational evaluator has a double task: determining appropriate criteria and applying them. It is a task in the best tradition of the most abstract theoretical science as well as the most practical applied science.